

Logo

Description automatically generated

**DataPLANT’s QuickStart on ARCs**

V1.3, May 2022

We are very happy that you chose our tools and infrastructure to create and share your own ARCs. In this QuickStart we focus on how to use the “ARC Commander” to store your data and “SWATE” to enrich it with metadata.

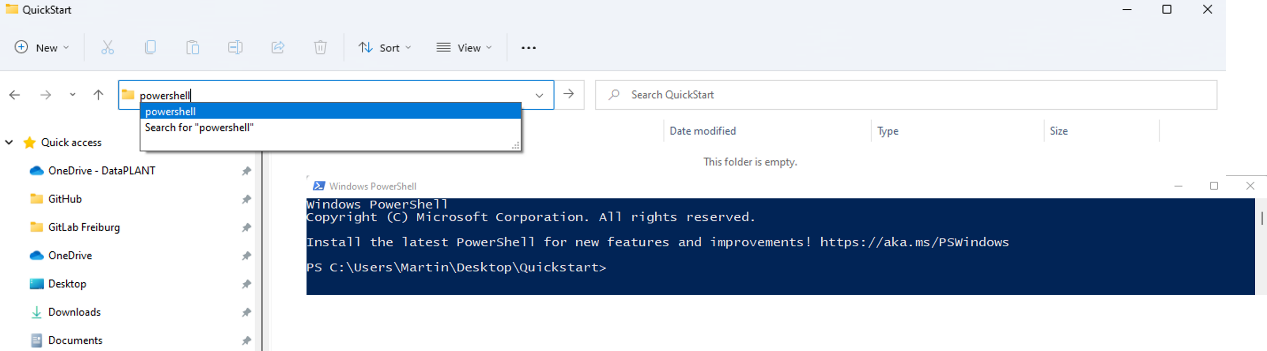
This document is work in progress. If you experience any inconsistencies, have questions or would like to suggest additions, please feel free to send a message to: [info@nfdi4plants.org](mailto:info@nfdi4plants.org) or open an issue at <https://github.com/nfdi4plants/quickstart>.

* [Environment and setup](#environment-and-setup)
  + [The command line](#the-command-line)
  + [Required software](#required-softwares)
* [ARC initialization](#arc-initialization)
* [Adding metadata](#adding-metadata)
  + [ISA investigation](#isa-investigation)
  + [ISA studies and assays](#isa-studies-and-assays)
* [Sharing your ARC](#sharing-your-arc)
  + [DataPLANT registration and access](#dataplant-registration-and-access)
  + [ARC synchronization](#arc-synchronization)
    - [Setting a git user](#setting-a-git-user)
  + [Invite collaborators](#invite-collaborators)
* [Data annotation](#data-annotation)
  + [SWATE](#swate)
    - [Customize your table by adding building blocks](#X56a0111bc6146e79b003d75f17f0e7aa12403a2)
    - [Use templates](#use-templates)
    - [Annotate your samples and data](#annotate-your-samples-and-data)
* [DataPLANT Support](#dataplant-support)
* [The Minimalist’s ARC-QuickStart](#the-minimalists-arc-quickstart)

## 

## Environment and setup

### The command line

* Most of this Quickstart (especially the section [ARC initialization](#arc-initialization)) is based on the command line (Windows: powershell; Linux and Mac: terminal).
* The following picture shows exemplarily how to open a powershell on windows by entering *powershell* into the explorer path:
* 
* Text formatted as code blocks represents commands to copy/paste into the command line:

echo "hello - I am a code block"

### 

### Required software

* Prerequisites for using the ARC Commander are [git](https://git-scm.com/downloads) and [git LFS](https://git-lfs.github.com/)

**Note:** If this is your first time using git on this computer, you need to set your git user name and email address. These are needed for displaying them on the git commits. You can update the settings with

git config --global user.name <your\_name>  
git config --global user.email <your\_email>

* Please download the latest version of the [ARC Commander](https://github.com/nfdi4plants/arcCommander/releases) for your operating system and install it according to [these instructions](https://github.com/nfdi4plants/arcCommander#install-and-start).
* Check if the ARC Commander is functional by displaying the ARC commander version and help menu:

arc --version  
arc --help

Text

Description automatically generated

## 

## ARC initialization

1. Create and navigate to a local folder, which you want to initialize as an ARC.

mkdir ~/Desktop/QuickStart;   
cd ~/Desktop/QuickStart

1. Initialize your ARC by executing

arc init

1. This will create the general ARC folder structure:

A picture containing graphical user interface

Description automatically generated

## 

## Adding metadata

### ISA investigation

The ISA investigation (-i) workbook allows you to record administrative metadata of your project. Add the isa.investigation.xlsx workbook including an identifier to your ARC with

arc i create -i QuickStartInvestigation

### ISA studies and assays

The ISA study (-s) and ISA assay (-a) workbooks allow you to annotate your experimental data.

1. Add an isa.study.xlsx workbook including an identifier to your ARC with

arc s add -s QuickStartStudy

1. Add an isa.assay.xlsx workbook including an identifier to your ARC with

arc a add -s QuickStartStudy -a QuickStartAssay

* The ARC Commander will add a subdirectories to the *studies* and *assays* folder. Your ARC should look similar to this now:

Graphical user interface, application

Description automatically generated

* **Note:** An assay must be linked to a study. If a study does not exist, it will be created automatically in this step.
* These steps can be repeated to add as many studies and assays as needed. Accordingly, more subdirectories will be added. Multiple assays can be grouped in a study when the same StudyIdentifier is used.

1. Place the data for each assay in the respective dataset folder.

## Sharing your ARC

### DataPLANT registration and access

In case you are not a member of DataPLANT yet, please visit <https://register.nfdi4plants.org> to register. Afterwards, you will be granted access to DataPLANT’s DataHUB, available under <https://git.nfdi4plants.org>. The DataHUB allows you to share your ARCs with registered lab or project partners.

Graphical user interface, text, application

Description automatically generated

After successful registration, create and set an access token for ARC Commander synchronization using

arc remote accesstoken get -s https://git.nfdi4plants.org

A window within your browser will open, asking for your DataPLANT Log In. In case you are already logged in, the browser will directly display a Success message to you:

Graphical user interface, text, application, chat or text message

Description automatically generated

### ARC synchronization

1. Synchronize your ARCs with the DataHUB using the command

arc sync

1. If you did not connect your local ARC with a remote one so far, you can specify the remote address with the flag -r followed by an URL, e.g.,

arc sync -r https://git.nfdi4plants.org/martinkuhl/QuickStart

1. In case you want to create a new remote repository at this URL, it needs to be assembled as the following example:

# https://git.nfdi4plants.org/<YourUserName>/<YourARC>

1. If no repository exists under the given URL, the ARC Commander will produce an error ensuring that you spelled the URL correctly. To force synchronization, use

arc sync -f

A picture containing chart

Description automatically generated

1. Check if the upload was successful by visiting your ARC at the respective URL in your browser.

Graphical user interface, text, application, email

Description automatically generated

**Note:** Alternatively, you can first create a new blank repository in the [DataHUB](https://git.nfdi4plants.org) by clicking “New project/repository” in the plus drop down menu of the navigation bar on top. Afterwards, you can sync your local ARC to the respective repository by adapting the URL to the newly generated one.

### Setting a git user

Some users might want to use different signatures for different repositories, e.g. for developing software on GitHub and working on ARCs on [DataPLANT’s DataHUB](htttps://git.nfdi4plants.org). Besides your global git configuration, you can store the information you want to use for editing ARCs within the ARC Commander config:

arc config set -g -n "general.gitname" -v "Name of choice"  
arc config set -g -n "general.gitemail" -v "Email of choice"

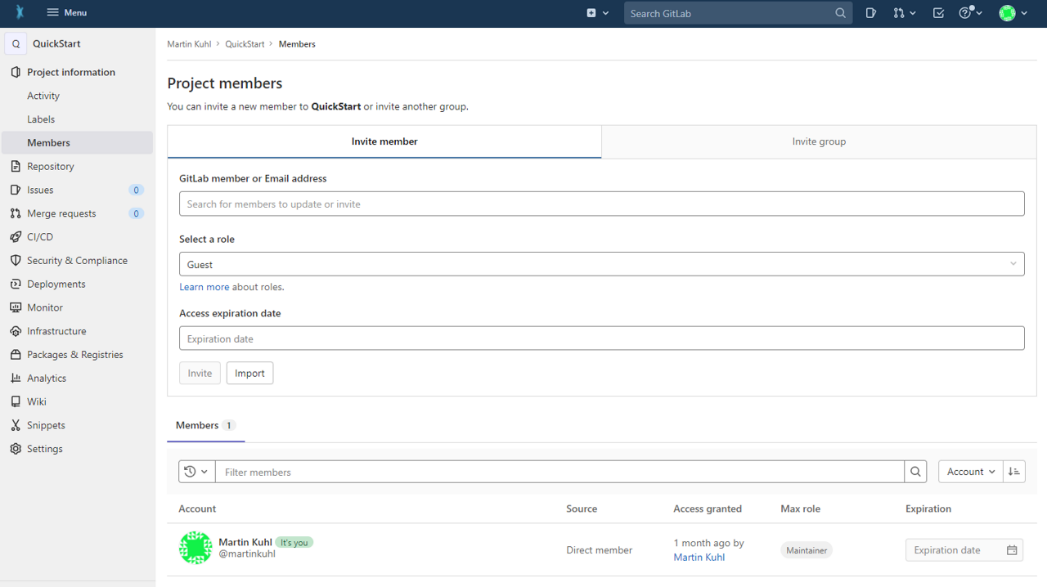
To transfer the information from the global ARC Commander config to the local git config of the ARC use

arc config setgituser

### Invite collaborators

You can invite lab-colleagues or project partners to join your ARC for collaboration. While inside your ARC on the DataHUB, click on *Project information -> Members* in the left navigation panel. Search for registered researchers and select a role for each individually. These roles come along with different rights.  
Briefly:

* *Guests:* Have the least rights. This is recommended for people you ask for consultancy.
* *Developers:* The choice for most people you want to invite to your ARC. Developers have read and write access, but cannot maintain the project on the DataHUB, e.g. inviting others.
* *Maintainers:* Gives the person the same rights as you have (except of removing you from your own project). This is recommended for inviting PIs or group leaders allowing them to add their group members for data upload or analysis to the project as well.



**Note:** A detailed usage instruction for the ARC Commander can be found [here](https://github.com/nfdi4plants/arcCommander/wiki/Detailed-usage-instruction).

## Data annotation

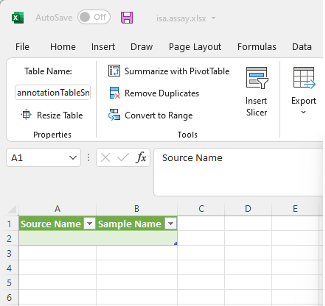
Your ARC should now contain one isa.investigation.xlsx and one or more isa.study.xlsx and isa.assay.xlsx file(s), respectively. Use the isa.study.xlsx to describe the characteristics of your samples, e.g. how you grew your plant, and isa.assay.xlsx to annotate the experimental analyses.

### SWATE

DataPLANT provides the Excel Add-In SWATE to support you in data annotation.

* Download and install the newest SWATE version according to [these instructions](https://github.com/nfdi4plants/Swate/wiki/docs01-installing-Swate#desktop-installation).
* In case you use an Excel version older than Excel 2019, please install [SWATE for Excel online](https://github.com/nfdi4plants/Swate/wiki/docs01-installing-Swate#quickstart).
* Use the *create annotation table* button in the yellow pop-up box (this only appears if you start SWATE on an Excel worksheet without an existing annotation table). An annotation table with the building blocks *Source Name* and *Sample Name* will be generated.

Graphical user interface, text, application, Excel

Description automatically generated

* Annotate your table with help of the [annotation principles](https://nfdi4plants.github.io/AnnotationPrinciples/).  
  Briefly:
  + *Characteristics* are used for study descriptions and describe inherent properties of the source material (e.g. a certain strain).
  + *Parameters* describe steps in your experimental workflow (e.g. an instrument model or a growth chamber), and
  + *Factors* describe independent variables that result in a specific output (e.g. the light intensity).
* The combination of ISA (Characteristics, Parameter, Factor) and a biological or technological ontology (e.g. temperature, strain, instrument model) gives the flexibility to display an ontology term, e.g. temperature, as a regular process parameter or as the factor your study is based on (Parameter [temperature] or Factor [temperature]).

#### Customize your table by adding building blocks

1. Choose the type of building block you want to add (A).
2. If you chose a descriptive building block type (building blocks besides Sample Name, Source Name, and Data File Name), use search field (B) to search for an Ontology Term. SWATE accesses the SwateDB with a list of established external ontologies designated suitable for use in plant science. In addition, we feature our own ontology NFDI4PSO to extend the DB with missing, but necessary terms.
3. If you want to add a building block with a unit, check box (C) and use search field (D) to look for a fitting unit term, e.g. degree Celsius as unit for Parameter [temperature].
4. If you could not find a fitting term, you can use the Advanced Term Search with the blue links above the *Add building block* button. If you still could not find a fitting term, use free text input.

Graphical user interface

Description automatically generated with medium confidence

1. For more information on customizing your annotation table click [here](https://github.com/nfdi4plants/Swate/wiki/Docs03-Building-Blocks).

#### Use templates

Alternatively, you can also use one of DataPLANT’s [SWATE templates](https://github.com/nfdi4plants/Swate/wiki/Docs05-Templates). You can find them under the *Protocol Insert* tab in SWATE.

Graphical user interface, text, application

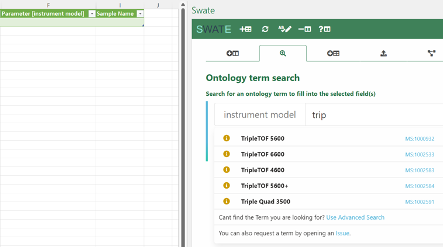
Description automatically generated

#### Annotate your samples and data

Fill the cells beneath each building block with ontology terms to note the respective *Characteristics, Parameter,* and *Factor* values of your experiment. Using the ontology term search function, you can fill multiple cells at once.

1. When *Use related term directed search* (A) is enabled, SWATE will suggest a selection of suitable terms within the ontology for the column header, e.g. *TripleTOF* *5600* for *instrument model.*
2. When term directed search (A) is disabled, SWATE will still suggest ontology terms, but without relation to the column header.
3. If you could not find a fitting term, use free text input.

Graphical user interface, text, application, Excel

Description automatically generated

**Note:** More information on how to use SWATE can be found [here](https://github.com/nfdi4plants/Swate/wiki/Docs05-Templates).

## DataPLANT Support

For further assistance, feel free to reach out via our [helpdesk](https://support.nfdi4plants.org) or by contacting us [directly](mailto:info@nfdi4plants.org).

## The Minimalist’s ARC-QuickStart

* ☒ You know how to use a command line
* ☒ You have created an ARC before
* ☒ The latest version of the [ARC Commander](https://github.com/nfdi4plants/arcCommander/releases) as well as [git](https://git-scm.com/downloads) and [git LFS](https://git-lfs.github.com/) are installed on your computer
* ☒ You have a [DataPLANT](https://register.nfdi4plants.org) account
* ☒ Your computer is linked to the [DataHUB](https://git.nfdi4plants.org) via an ssh key or a personal access token.

Voila! You are ready to follow these few steps to create a minimal ARC sharable via DataPLANT’s DataHUB:

1. Visit the [DataHUB](https://git.nfdi4plants.org), create a new repository and copy the URL to your clipboard.
2. Replace the <variables> in the following code block with your information and execute it in your command line.

# Create and navigate to your ARC folder  
mkdir <YourARC>  
cd <YourARC>  
  
# Setup the ARC structure with one study and one assay  
arc init  
arc i create -i <YourInvestigation>  
arc a add -s <YourStudy> -a <YourAssay>  
arc sync -f -r https://git.nfdi4plants.org/<YourUserName>/<YourARC> -m "initialize ARC structure"